Random matrix-based dimensionality reduction for bioluminescence tomography reconstruction

Styles, I.B.; Basevi, H.R.A.; Guggenheim, J.A.; Dehghani, H.

DOI:
10.1117/12.2033299

Document Version
Peer reviewed version

Citation for published version (Harvard):
Styles, IB, Basevi, HRA, Guggenheim, JA & Dehghani, H 2013, Random matrix-based dimensionality reduction for bioluminescence tomography reconstruction. in *Progress in Biomedical Optics and Imaging - Proceedings of SPIE.* vol. 8799. https://doi.org/10.1117/12.2033299

Random Matrix-based Dimensionality Reduction for Bioluminescence Tomography Reconstruction

Iain B. Styles[a], Hector R.A. Basevi[a,b], James A. Guggenheim[a,b], and Hamid Dehghani[a]

[a]School of Computer Science, University of Birmingham, Edgbaston, Birmingham, B15 2TT,United Kingdom

[b]PSIBS, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

ABSTRACT

We show how a random matrix can be used to reduce the dimensionality of the bioluminescence tomography reconstruction problem. A randomised low-rank approximation for the sensitivity matrix is computed, and we show how this can be used to reconstruct the bioluminescence source distribution on a randomised basis for the mesh nodes. The distribution on the original mesh can be found easily via a simple matrix multiplication. The majority of the computation required can be performed in advance of the reconstruction, and the reconstruction time itself is of the order milliseconds. This could allow for high frame rate real-time reconstructions to be performed.

Keywords: bioluminescence tomography, inverse problems, reconstruction

1. INTRODUCTION

Reconstruction in bioluminescence tomography involves the computation of the internal distribution of light sources within a subject (usually a small animal) from fluence measurements made at the subject's surface. It is a widely used tool in preclinical molecular imaging, where specific cells of interest (such as cancer cells) are tagged with genes for the production of bioluminescent compounds such as firefly luciferase. The reconstruction process requires the strength of emission throughout the interior of the subject (usually defined on a finite-element mesh) to be calculated from measurements taken from around the surface. Since the surface fluence is linear with regard to the interior sources, the reconstruction problem amounts to solving a system of linear equations relating the strength of measurements at the surface of the object to the luminescent source distribution within the object, via a physical model matrix describing the response of the measurements to the source distribution. Such sets of linear equations are superficially simple, but it is difficult to develop effective solution schemes because the number of measurements is normally far less than the number of unknowns that we wish to reconstruct: the inverse problem is under-determined, ill-conditioned and ill-posed, with many possible solutions. A number of approaches have been suggested for the efficient solution of such problems including standard methods that find the solution with the shortest Euclidean length (L_2-norm) and compressive sensing-inspired methods that employ L_1 regularisation to find the sparsest solution.[1] The disadvantage of these methods is that they are usually iterative algorithms that are time-consuming.

We propose a method for the dimensionality reduction of the inverse problem that allows solutions to be found at low computational cost, and with high speed. We generate an approximate basis for the mesh nodes such that each basis vector is randomly distributed over all of the real mesh nodes, and reconstruct the bioluminescence distribution on this lower-dimensional representation of the mesh. A simple transformation then allows the solution to be mapped onto the real mesh nodes. We demonstrate the method on simulated data generated using the NIRFAST modelling software[2] from the mouse atlas model "Digimouse"[3] and compare with a recently develop L_1-regularised method that has previously been shown to be effective on similar problems.[1] The proposed method is shown to be effective in recovering source distribution with source locations identified to within a similar level of accuracy as the L_1 method, but the computation time is greatly reduced. In particular, the majority of the computations can be pre-computed at the time the model matrix J is generated, and the reconstruction time itself takes less than a second.

Corresponding author: Iain Styles, I.B.Styles@cs.bham.ac.uk

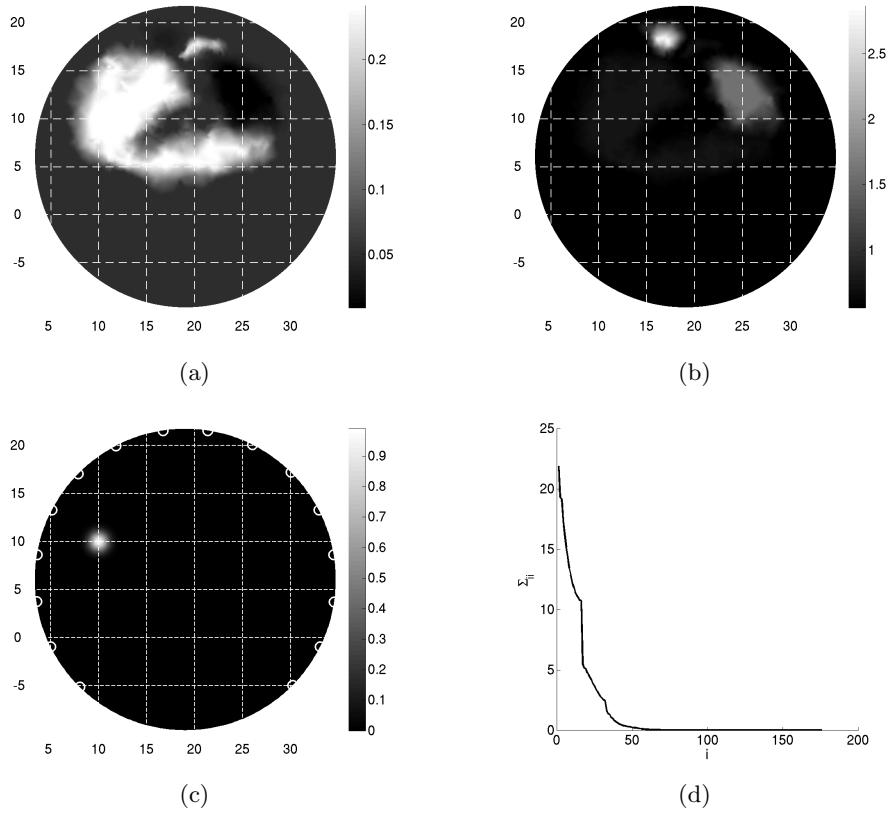

Figure 1. Properties of simulated atlas data: a) μ_a at 560nm; b) μ_s at 560nm; c) Bioluminescence source distribution, with detector positions indicated by white circles at the boundary. Gridlines at 5mm spacing are shown for reference. d) Ordered singular values of the sensitivity matrix $\mathbf{J}$.

2. METHODS

We consider a simulated dataset consisting of a single bioluminescent source embedded in a 2d slice derived from a commonly used mouse atlas.[3] The slice contains 8358 FEM mesh nodes with 176 measurements available from 16 detectors placed around the boundary of the slice, each of which is sensitive to 11 discrete wavelengths. The reconstruction problem is therefore to calculate the bioluminescence source distribution over 8358 mesh nodes from 176 measurements. The target source distribution is shown in Figure 1, together with the optical properties at 560nm. Simulated measurements were computed, given the source distribution and the optical properties, using the finite-element tissue optics modelling software NIRFAST.[2]

In order to reconstruct the source distribution on the mesh nodes from the measurements, we must solve the set of linear equations

$$\mathbf{y} = \mathbf{J}\mathbf{x} \tag{1}$$

where the coefficients of $\mathbf{y}$ are the values of the 176 measurements, the coefficient of $\mathbf{x}$ are the strengths of the bioluminescent sources at the nodes of the FEM mesh, and the coefficients of $\mathbf{J}$, J_{ij} describe the response of detector i to a source at position j.

Although this problem appears to be, and indeed is highly under-determined, it is useful to note that $\mathbf{J}$ must be highly rank-deficient. From the fundamental theorem of linear algebra, it is known that $\text{rank}(\mathbf{J}) = \text{rank}(\mathbf{J}^{\text{T}})$, and it therefore follows that maximum possible rank of $\mathbf{J}$ is equal to its smallest dimension (in this case, the number of measurements–176). It therefore follows that the columns of $\mathbf{J}$ (the "Jacobian"), corresponding to the nodes of the FEM mesh, must contain linear dependencies. Extracting these dependencies requires some type of rank-revealing factorisation such as singular value decomposition (SVD) or QR factorisation to be performed.

This can be very resource-intensive due to the large matrices that must be constructed. For the current problem, the SVD can be done with reasonable computational effort, and the singular values are shown in Figure 1(d). $\mathbf{J}$ is, in fact, full-rank, but the singular values decay quite rapidly and many are very small which implies that the matrix could be replaced by a low-rank approximation. This can be achieved using rank-revealing transformations such as the SVD, but only if the SVD can be computed easily.

An increasingly popular technique for the analysis of large structured datasets is the random projection. It has been established that projecting high-dimensional datasets onto a relatively small number of random vectors preserves, with high probability, the essential features of the data (such as angles between vectors, and distances between samples).[4] Although the simplest forms of the random projection are not directly applicable to the current problem, a variant proposed recently by Halko et al[5] has shown how random projections can be used to generate approximate low-rank representations of large rank-deficient matrices, and has recently found application in the analysis of large spectroscopic datasets.[6] We draw on their work to efficiently generate a low-rank approximation for $\mathbf{J}$. The following procedure is adapted from Algorithm 4.1 therein.

- Given measurements $\mathbf{y}_{m\times 1}$, Jacobian $\mathbf{J}_{m\times n}$ and source distribution $\mathbf{x}_{n\times 1}$
  - Draw a real-valued random matrix $\mathbf{\Omega}_{n\times l}$ with $\Omega_{ij} \sim \mathcal{N}(0,1)$.
  - Form $\mathbf{Y}_{n\times l} = (\mathbf{J}^{\mathrm{T}})_{n\times m}\mathbf{\Omega}_{n\times l}$
  - Generate a basis for $\mathbf{Y}$ by QR-factorisation: $\mathbf{Y}_{n\times l}$=$\mathbf{Q}_{n\times l}\mathbf{R}_{l\times l}$, where $\mathbf{R}$ is upper-triangular and the columns of $\mathbf{Q}$ form a basis for the rows of $\mathbf{Y}$.
- If $l \geq \mathrm{rank}(\mathbf{J})$ then $\|\mathbf{J} - \mathbf{J}\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\| < \epsilon$ for small ϵ (to high probability), and the low-rank (l) factorisation $\mathbf{J}\mathbf{Q}\mathbf{Q}^{\mathrm{T}}$ can be used as an approximation for $\mathbf{J}$.

The solution to Eq. (1) can then be formulated as a pair of sub-problems. In the first, we replace $\mathbf{J}$ by its approximate low-rank factorisation to obtain

$$\mathbf{y} \approx \mathbf{J}\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{x} = \tilde{\mathbf{J}}\tilde{\mathbf{x}} \tag{2}$$

where $\tilde{\mathbf{J}} = \mathbf{J}\mathbf{Q}$ and $\tilde{\mathbf{x}} = \mathbf{Q}^{\mathrm{T}}\mathbf{x}$. This must be solved for $\tilde{\mathbf{x}}$. Since $\tilde{\mathbf{J}}_{n\times l}$ is, by design, full or nearly full-rank, the solution of this problem can be trivially found using any efficient solver. In this work, the MATLAB operation `mldivide` was used.

The second sub-problem is to solve $\tilde{\mathbf{x}} = \mathbf{Q}^{\mathrm{T}}\mathbf{x}$ for $\mathbf{x}$. This is an under-determined problem that would normally be at least as difficult as the original problem. However, the factorised form of $\mathbf{J}$ allows a simple approach. Since $\mathbf{J} \approx \mathbf{J}\mathbf{Q}\mathbf{Q}^{\mathrm{T}}$ and therefore $\mathbf{J}\mathbf{x} \approx \mathbf{J}\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{x}\ \forall\mathbf{x}$, it follows that $\mathbf{x} \approx \mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{x}$, or $\mathbf{x} \approx \mathbf{Q}\tilde{\mathbf{x}}$.

It is important to note that the quality of the factorisation is dependent on the random matrix $\mathbf{\Omega}$ being close to its expectation value. Normally this will be ensured by choosing small ϵ to ensure a good approximation, but it may be that there are some pathological cases where this condition is met, but that the approximation breaks down for some particular $\mathbf{x}$. In order to guard against this we generate multiple randomised bases $\mathbf{Q}^{(k)}$ for $k = 1 \rightarrow K$ and compute $\mathbf{x} = \frac{1}{K}\sum_{i=1}^{K}\mathbf{x}^{(k)}$, where $\mathbf{x}^{(k)}$ is the bioluminescence distribution recovered using the corresponding basis set.

These methods were applied to simulated measurements generated from the Digimouse phantom using the bioluminescence distribution shown in Figure 1(c). This was done by computing the sensitivity matrix using the NIRFAST software and then applying this to the source distribution vector. The measurements and Jacobian were then scaled such that the maximum measurement was equal to 90% of the maximum signal of a 16-bit detector. The reconstruction process was applied to these measurements with zero noise added, and with shot noise added: $y_i \mapsto y_i + \mathcal{N}(0, y_i)$.

3. RESULTS AND DISCUSSION

Bioluminescence source distributions were recovered from both noise-free and noisy simulated measurements. Results were averaged over $K = 20$ random bases and where noise was added, reconstructions were also performed on 20 realisations of noise. In both noise-free and noisy cases, the rank of the approximation was varied by changing the number (l) of random projection vectors used. As the original Jacobian is full-rank (176), the rank of the approximation was varied from 25 to 250 in order to investigate the transition between low-rank and full-rank approximations. The averaged (over the bases) reconstruction results for the noise-free case are shown in Figure 2, and the results for the noisy case for a single realisation of the noise are shown in Figure 3. For display purposes, a small number ($< 0.1\%$) of data points with negative values were set to zero. These were exclusively located at nodes located near the detectors at the boundary, a known area of difficulty in tomographic reconstruction problems.

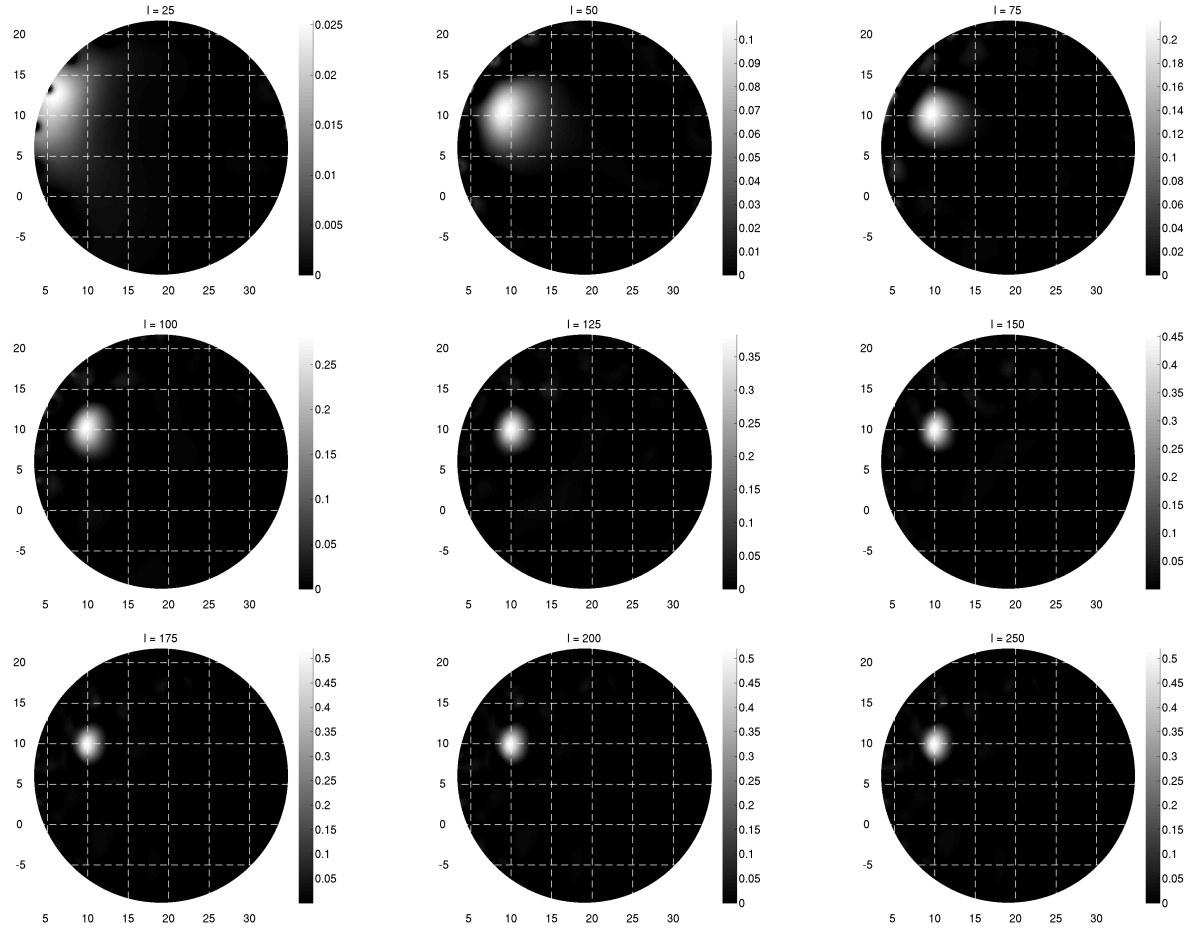

Figure 2. Reconstructions from simulated noise-free measurements using approximations of different rank. Results shown are averaged over 20 random bases. A small number of negative points located near to the detectors have been set to zero.

In the results from noise-free data (Figure 2), we observe that the reconstructions are identical when the rank of the approximation is higher than the rank of the data ($l = 200, 250$). This is expected as the dimensionality of the approximate representation is greater than the rank of the original matrix. When the rank of the approximation drops below that of the Jacobian, the quality of the reconstruction degrades gradually, but visibly, by visual inspection, with the peak location drifting to the upper left of the reconstructed images, accompanied by a noticeable broadening of the peak width. The reconstructions remain plausibly representative of the known

ground truth down to $l = 50$. This can be ascribed to the behaviour of the singular values shown in Figure 1d, where it can be seen that it is at around 50 that the singular values begin to approach zero.

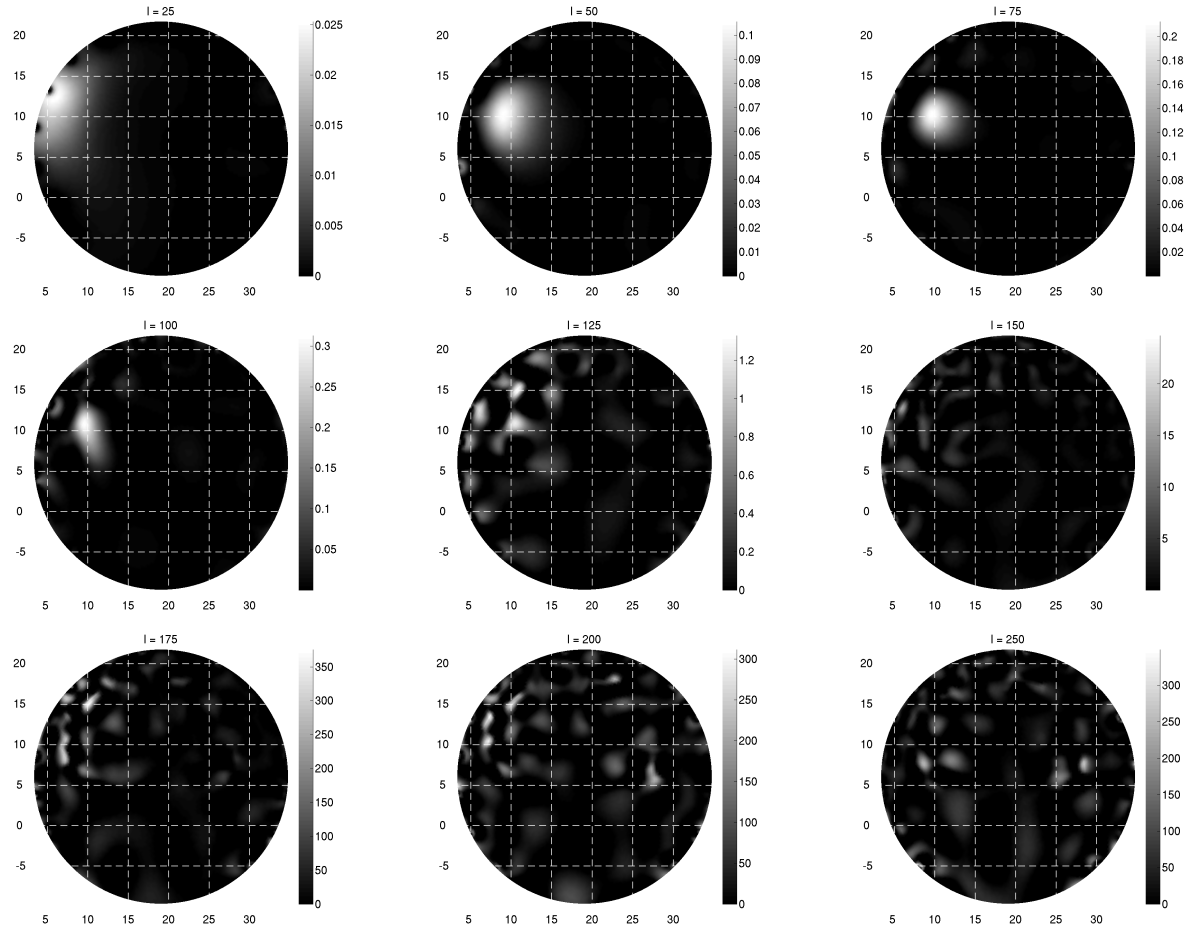

Figure 3. Reconstructions from simulated measurements with added shot noise using approximations of different rank. Results shown are for a single realisation of noise, and averaged over 20 random bases. A small number of negative points located near to the detectors have been set to zero.

The reconstructions from noisy data (Figure 3) show a significantly different pattern. For higher-rank approximations, it is not possible to determine the position of the bioluminescence peak. However, with the Jacobian projected onto 50, 75 and (to a lesser extent) 100 projections, the peak is clearly visible, and is accurately located. This is a striking result that requires further analysis. The higher rank approximations, which should in principle be more representative of the Jacobian, perform qualitatively worse than lower-rank approximations. Here, it is worth revisiting the SVD (Figure 1(d)) again. We note that 50, 75 and 100 projections are still representative of the majority of the matrix's structure, but since there are fewer free parameters to fit than in the higher-dimensional cases, over-fitting is less problematic. The source of this is the first part of the solution scheme where $\tilde{\mathbf{x}}$ is computed. When $l < m$, the problem is *overdetermined*, and hence the solution at this stage is a smooth best-fit to the points. For the higher-rank cases, a closer fit is found, but this is not desirable due to the noise: oversampling leads to over-fitting. It therefore seems to be optimal that the number of projections used should be as low as possible, whilst ensuring that the majority of the information is preserved.

Two performance measures have been calculated for these results. The peak position, defined as the coordinate of the mesh node with the highest value, has been found where possible (it was not possible to identify a main peak in the higher-rank noisy reconstructions), and the displacement from the known peak location calculated.

These results are given in Table 1

Rank	25	50	75	100	125	150	175	200	250
Noise-free	4.59	0.90	0.61	0.28	0.10	0.10	0.21	0.21	0.21
Noisy	4.62±3.08	1.12±0.51	0.60±0.10	1.11±6.67	–	–	–	–	–

Table 1. The distance between the main peak and the known target. Mean and standard deviation over 20 realisations of noise are quoted for the reconstructions with added noise.

In order to compare these results with an established algorithm, we performed the same reconstructions using an L_1 regularised algorithm[1] which has been shown to be effective on both simulated and experimental data containing either single or multiple sources. This algorithm was able to recover the peak to a typical accuracy of 0.34±0.0mm in both noise-free and noisy cases (20 realisations of noise). The reconstructions from this method are shown in Figure 4. The proposed method is therefore comparable in peak localisation accuracy to established methods, although some caution is required as these values are somewhat influenced by the discrete nature of the mesh which limits the accuracy to which the peak position can be determined. Qualitatively we observe that the L_1-minimal solutions are more tightly localised than the proposed method's, which is expected because the L_1 constraint enforces a sparse solution. With regard to the intensity of the recovered distributions, neither method recovers the peak height exactly, although the L_1 method is superior in this regard. The solution generated by the randomised method is believed to be the L_2 minimal solution (no formal proof has been found), which tends to enforce smoothness on the solution which inevitably broadens and lowers the peak height.

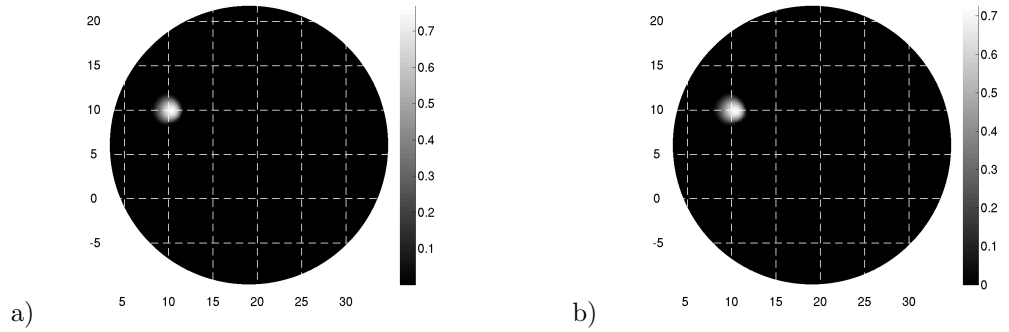

Figure 4. Reconstructions performed using an L_1 regularised algorithm.[1] a) Reconstruction from noise-free measurements; b) Reconstruction from measurements with added shot noise.

The second aspect of algorithm performance that is of importance is the execution time of the algorithm. In Figure 5 the execution time on a computer with a dual-core Intel Xeon 2.4GHz processor is shown. The execution time is broken down into two parts: (i) the computation of $\mathbf{Q}$ and $\tilde{\mathbf{J}}$; (ii) solving for $\tilde{\mathbf{x}}$ and $\mathbf{x}$. The times are given for computations using 20 random bases over which the results are averaged, in order to ensure stability and reproducibility of the results.

It is apparent from these results that the majority of the computational effort is in the generation of the basis, whilst the reconstruction itself is very fast (fractions of a second). This is significant: the basis generation need only be done once and can be done off-line (i.e. before the reconstruction is required), for example, at the same time as the sensitivity matrix is calculated. At the time of reconstruction, only the two solution stages need be performed, and this can be done extremely quickly (for example, with 100 projection dimensions, the reconstruction takes 0.25 seconds).

4. CONCLUSIONS

We have investigated the use of random projections to reduce the dimensionality of the reconstruction problem in bioluminescence tomography. Our results indicate that a low-rank approximation of the sensitivity matrix

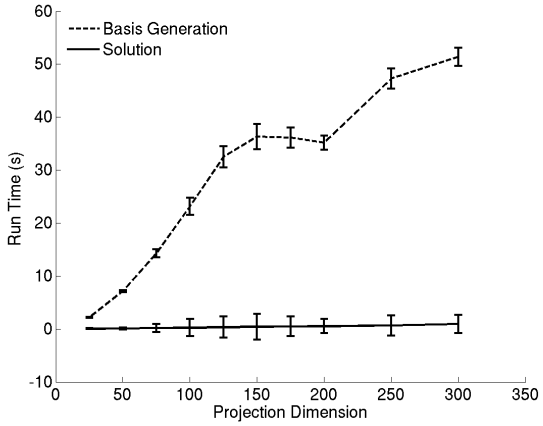

Figure 5. The execution time of the randomised reconstruction algorithm separated into dimensionality reduction and solution stages. The times shown are those required for 20 reconstructions so that the results can be averaged.

can be found through randomised methods, and that this can be used to obtain localisation accuracy that is similar to that achievable by iterative methods. It is believed that the proposed methods yields the L_2–minimal solution to the problem with the consequence that the recovered peak is lower in height and more spread than the target, due to the known smooth nature of such solutions. We also observed that when noise is present in the measurements, some degree of under-sampling is desirable. It appears that there is an optimal "window" between the point at which the singular values start to become significant and the true rank of the matrix, in which the number of samples should lie such that the essential character of the matrix is preserved, but over-fitting of the noise is avoided. An analysis of the computation time revealed that the bulk of the computation is required in the basis generation stage, and it was noted that this can be done at the same time as the sensitivity matrix is computed, i.e. in advance of the reconstruction itself. The reconstructions themselves can be performed in fractions of a second. This opens up the possibility of high-frame-rate, real-time reconstruction.

5. ACKNOWLEDGEMENTS

The authors acknowledge useful discussions with Bob Durrant and Ata Kaban. H. R. A. B. and J. A. G are grateful for financial support from the EPSRC under grant EP/F50053X/1 as part of the PSIBS Doctoral Training Centre.

REFERENCES

1. H. R. A. Basevi, K. M. Tichauer, F. Leblond, H. Dehghani, J. A. Guggenheim, R. W. Holt, and I. B. Styles, "Compressive sensing based reconstruction in bioluminescence tomography improves image resolution and robustness to noise," Biomed. Opt. Express **3**, 2131–2141 (2012).
2. H. Dehghani, M. E. Eames, P. K. Yalavarthy, S. C. Davis, S. Srinivasan, C. M. Carpenter, B. W. Pogue, and K. D. Paulsen, "Near infrared optical tomography using nirfast: Algorithm for numerical model and image reconstruction," Communications in Numerical Methods in Engineering **25**, 711–732 (2009).
3. B. Dogdas, D. Stout, A. Chatziioannou, and R. Leahy, "Digimouse: a 3d whole body mouse atlas from ct and cryosection data," Physics in medicine and biology **52**, 577 (2007).
4. E. Bingham and H. Mannila, "Random projection in dimensionality reduction: applications to image and text data," in "Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining," (ACM, 2001), pp. 245–250.
5. N. Halko, P. G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," SIAM Review **53**, 217–288 (2011).
6. A. D. Palmer, J. Bunch, and I. B. Styles, "Randomized approximation methods for the efficient compression and analysis of hyperspectral data," Analytical Chemistry **85**, 5078–5086 (2013).